

機械学習を用いた Knight- Amazons の対戦プログラムの実装

平賀 太竣 † 盧 金陽 † Fazekas Szilard Zsolt † 山村 明弘 †

† 秋田大学理工学研究科

1 はじめに

Knight-Amazons[3]とは加藤光らが Amazons[7]のルールを一部変更して作ったオリジナルの二人対戦ゲームである。Knight-Amazons には決まった盤のサイズ、駒の個数、配置はないので、今回の実験では次節の図1のように定義する。本研究では、モンテカルロ木探索プログラムを用いて学習データを作成する。そのデータを学習したニューラルネットワークとモンテカルロ木探索プログラムの性能を比較し、人間の対戦データなしでプログラムが強くなるかを検証した。

2 Knight-Amazons

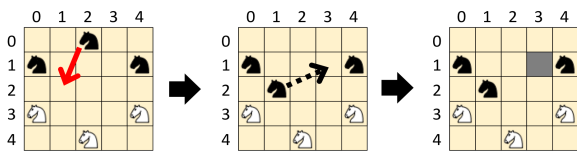


図 1: Knight-Amazons のゲーム例

ゲームに使用する駒の移動範囲はチェスの Knight の動きと同じである。プレイヤーは自分の手番になったら、自分の色の駒から一つ選択し、移動範囲に従って駒が配置されていない選択可能なマスから一つを選択し、駒を移動させる。その後、移動したマスから同じようにマス一つを選択し、その選択したマスを潰す。潰したマスはそのゲーム中は選択不可能なマスとなる。マスを潰したら手番は次のプレイヤーに移る。この動作を交互に行い、先に駒を動かすことができなくなったプレイヤーが敗者となる。図1に黒の手番での一連の手順を示している。実戦矢印が駒の動きを表し、点線矢印が潰すマスの指定した位置を表している。灰色のマスが選択不可能なマスである。

3 モンテカルロ木探索

モンテカルロ木探索とは、ある局面からランダムにゲームを終局までシミュレートしてその局面の価値を

推定します。この1回のシミュレートはプレイアウトと呼ばれていて、このプレイアウトを何度も繰り返し、勝率の大きい着手を見つけ、選択することができる。モンテカルロ木探索はこの原理から、決まった評価関数を用いる必要がないので、評価の基準が複雑な Knight-Amazons のようなゲームの状態を評価することができる。今回の実験では代表的なモンテカルロ木探索プログラムである UCT プログラム [6] を実装する。

4 実験概要

UCT プログラムとニューラルネットワークを対戦させる実験を行う。実験に使うニューラルネットワークは「40000 個のデータを学習したノード数 128 個の 1～5 層の全結合層のネットワーク」と「100000 個のデータを学習したノード数 128 個の 1～5 層の全結合層のネットワーク」と「ノード数 256 個の 1～5 層の全結合層のネットワーク」の計 15 種類を使う。またこれらのネットワークは、5x5 の盤面を一行に並べ替えた長さ 25 の一次元 Numpy 配列を入力データとして、選択する着手を 1、他を 0 とするような長さ 424 の一次元 Numpy 配列に変換して出力データとするようなネットワークである。これらのネットワークの共通パラメータを表1に示す。

活性化関数	ReLU 関数 (中間層), Softmax 関数 (出力層)		
最適化手法	Adam	損失関数	交差エントロピー誤差
学習率	0.001	ドロップアウト率	0.5
エポック数	100	バッチサイズ	32

表 1: 共通パラメータ

対戦に使う UCT プログラムはプレイアウト回数を 1000 回、設定する定数 C が全て 1.4 とする。ここで言っている定数 C は「探索されていない局面に費やす時間」と「有望な局面に費やす時間」の二つの間の優先度のバランスを表すパラメータである。実験は先手・後手をそれぞれ各 1000 回ずつニューラルネットワークと UCT プログラムを対戦させ、ニューラルネットワークの UCT プログラムに対する勝利数を記録する。

Knight-Amazons battle program using machine learning
 †Takatoshi Hiraga Jinyang Lu Fazekas Szilard Zsolt Akihiro Yamamura
 †Akita University

5 結果

実験のニューラルネットワークが先手の場合の勝利数を図2に示し、ニューラルネットワークが後手の場合の勝利数を図3に示す。図2、図3の縦軸は勝利数を表しており、横軸はネットワークの層の数である。図2、図3に表示されている凡例 A、B、C はそれぞれ、「学習データ 40000 個、ノード数 128 個」、「学習データ 100000 個、ノード数 128 個」、「学習データ 100000 個、ノード数 256 個」である。実験結果より、先手の場合であれば学習データ 40000 個のノード数 128 個の 3 層ネットワークと 100000 個のノード数 256 個の 4 層ネットワークが UCT プログラムに対して 500 回に近い勝利数を取めている。しかし、その二つのネットワークは後手の場合では共に勝利数が 400 回を下回っており、先手の場合の結果と合わせて考えると UCT プログラムと実力が同等とは言えない結果となった。仮説として先手が後手よりも有利のため、後手の場合の勝利数が下がると考えていたが、実験の結果を確認すると、先手の場合の勝利数よりも後手の場合の勝利数が多いネットワークも存在しており、先手・後手に有利不利が必ずあるとは言い切れないことがわかった。次に凡例 A、B、C の勝利数の違いについて考える。三種類の凡例を比較した時、全体的に凡例 A が勝利数は多く、学習データの個数が多くても勝利数にはそこまで影響はないように考えられた。また、先手の勝利数の結果での凡例 A、B の二つの最大勝利数と凡例 C での最大勝利数のそれぞれのネットワークの層の数を確認すると、凡例 C だけが 4 層のネットワークであった。これを学習データの数が増加したことで、対応する為に必要なネットワーク内のパラメータの数も増加したからと考えていたが、後手の結果での凡例 A は学習データの数が他の二つの凡例よりも少ないのにも関わらず、他の二つよりも層の数が多い方の勝率がよかった。

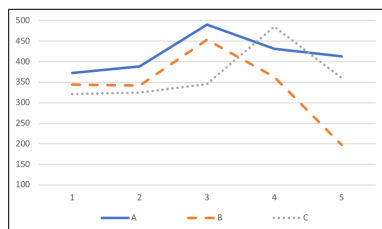


図2: ニューラルネットワークの勝利数 (先手)

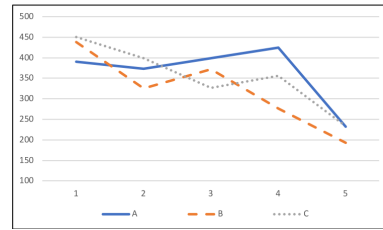


図3: ニューラルネットワークの勝利数 (後手)

6 まとめ

今回の実験では UCT プログラムよりも強い対戦プログラムを作ることはできなかったが、機械は人間の対戦データがなかったとしてもゲームのルールを覚え、ある程度は強くなることがわかった。今後の課題として、ネットワークの学習の理解を深め、学習データの数に対してより最適なネットワークの構造を見つけ出し、より強い対戦プログラムを作成したい。

参考文献

- [1] Max Pumperla, Kevin Ferguson, 山岡忠夫, 『囲碁ディープラーニングプログラミング』, マイナビ出版 (2019).
- [2] 美添一樹, 山下宏, 『コンピュータ囲碁 —モンテカルロ法の理論と実践—』, 共立出版 (2012).
- [3] Kato, H., Takaya, M., Yamamura, A.: Analysis of a Monte Carlo Tree Search in Knight- Amazons, *Procedia Computer Science*, 62 pp. 31-38 (2015).
- [4] Kato, H., Fazekas, S., Takaya, M., Yamamura, A.: Comparative Study of Monte-Carlo Tree Search and Alpha-Beta Pruning in Amazons. *Information and Communication Technology Vol 9357*, pp. 139-148 (2015).
- [5] ———: 『Amazons と Knight-Amazons の対戦プログラムの検討』 (2016).
- [6] Kocsis, L., Szepesvári, C.: Bandit based Monte Carlo planning, *17th European Conf. on Machine Learning (ECML 2006)*, pp. 282-293 (2006).
- [7] Zamkuskas, W.: Amazons, <<http://www.chessvariants.org/other.dir/amazons.html>> (参照 2020-1-10).